On the List-Decodability of Random Linear Codes

Venkatesan Guruswami* Johan Håstad† Swastik Kopparty‡

January 9, 2010

Abstract

For every fixed finite field $\mathbb{F}_q$, $p \in (0, 1-1/q)$ and $\varepsilon > 0$, we prove that with high probability a random subspace C of $\mathbb{F}_q^n$ of dimension $(1-H_q(p)-\varepsilon)n$ has the property that every Hamming ball of radius pn has at most $O(1/\varepsilon)$ codewords.

This answers a basic open question concerning the list-decodability of linear codes, showing that a list size of $O(1/\varepsilon)$ suffices to have rate within ε of the "capacity" $1-H_q(p)$. This matches up to constant factors the list-size achieved by general random codes, and gives an exponential improvement over the best previously known list-size bound of $q^{O(1/\varepsilon)}$.

The main technical ingredient in our proof is a strong upper bound on the probability that ℓ random vectors chosen from a Hamming ball centered at the origin have too many (more than $\Theta(\ell)$) vectors from their linear span also belong to the ball.

*Computer Science Department, Carnegie Mellon University. guruswami@cmu.edu. Supported in part by NSF CCF 0953155 and a Packard Fellowship.

†School of Computer Science and Communication, KTH. johanh@csc.kth.se. Research supported by ERC grant 226203.

‡CSAIL, MIT. swastik@mit.edu. Work was partially done while the author was an intern at Microsoft Research, New England.

1 Introduction

One of the central problems in coding theory is to understand the trade-off between the redundancy built into codewords (aka the rate of the code) and the fraction of errors the code enables correcting. Suppose we are interested in codes over the binary alphabet (for concreteness) that enable recovery of the correct codeword $c \in \{0,1\}^n$ from any noisy received word r that differs from c in at most pn locations. For each c, there are about $\binom{n}{pn} \approx 2^{H(p)n}$ such possible received words r, where $H(x) = -x\log_2 x - (1-x)\log_2(1-x)$ stands for the binary entropy function. Now for each such r, the error-recovery procedure must identify c as a possible choice for the true codeword. In fact, even if the errors are randomly distributed and not worst-case, the algorithm must identify c as a candidate codeword for most of these $2^{H(p)n}$ received words, if we seek a low decoding error probability. This implies that there can be at most $\approx 2^{(1-H(p))n}$ codewords, or equivalently the largest rate R of the code one can hope for is $1 - H(p)$.

If we could pack about $2^{(1-H(p))n}$ pairwise disjoint Hamming balls of radius pn in $\{0,1\}^n$, then one can achieve a rate approaching $1 - H(p)$ while guaranteeing correct and unambiguous recovery of the codeword from an arbitrary fraction p of errors. Unfortunately, it is well known that such an asymptotic "perfect packing" of Hamming balls in $\{0,1\}^n$ does not exist, and the largest size of such a packing is at most $2^{(\alpha(p)+o(1))n}$ for $\alpha(p) < 1 - H(p)$ (in fact $\alpha(p) = 0$ for $p \geqslant 1/4$). Nevertheless, it turns out that it is possible to pack $2^{(1-H(p)-\varepsilon)n}$ such Hamming balls such that no $O(1/\varepsilon)$ of them intersect at a point, for any $\varepsilon > 0$. In fact a random packing has such a property with high probability.

List Decoding. This fact implies that it is possible to achieve rate approaching the optimal $1 - H(p)$ bound for correcting a fraction p of *worst-case* errors in a model called *list decoding*. List decoding, which was introduced independently by Elias and Wonzencraft in the 1950s [Eli57, Woz58], is an error-recovery model where the decoder is allowed to output a small list of candidate codewords that must include all codewords within Hamming distance pn of the received word. Note that if at most pn errors occur, the list decoder's output will include the correct codeword. In addition to the rate R of the code and the error fraction p, list decoding has an important third parameter, the "list-size," which is the largest number L of codewords the decoder is allowed to output on any received word. The list-size thus bounds the maximum ambiguity in the output of the decoder.

For codes over an alphabet of size q, all the above statements hold with $H(p)$ replaced by $H_q(p)$, where $H_q(x) = x\log_q(q-1) - x\log_q x - (1-x)\log_q(1-x)$ is the q-ary entropy function.

Definition 1 (Combinatorial list decodability property)**.** *Let Σ be a finite alphabet of size q, $L \geqslant 1$ an integer, and $p \in (0, 1-1/q)$. A code $C \subseteq \Sigma^n$ is said to be (p, L)-list-decodable, if for every $x \in \Sigma^n$, there are at most L codewords of C that are at Hamming distance pn or less from x. Formally, $|B_n^q(x,p) \cap C| \leqslant L$ for every x, where $B_n^q(x,p) \subseteq \Sigma^n$ is the ball of radius pn centered at $x \in \{0,1\}^n$.*

We restrict $p < 1-1/q$ in the above definition since a random string differs from each codeword in at most a fraction $1 - 1/q$ of positions, and so over alphabet size q decoding from a fraction $1 - 1/q$ or more errors is impossible (except for trivial codes).

Combinatorics of list decoding. A fundamental question in list decoding is to understand the trade-off between rate, error-fraction, and list-size. For example, what list-size suffices if we

want codes of rate within ε of the optimal $1 - H_q(p)$ bound? That is, if we define $L_{q,p}(\varepsilon)$ to be the minimum integer L for which there are q-ary (p, L)-list-decodable codes of rate at least $1 - H_q(p) - \varepsilon$ for infinitely many lengths n, how does $L_{q,p}(\varepsilon)$ behave for small ε (as we keep the alphabet size q and $p \in (0, 1 - 1/q)$ fixed)?

It is known that unbounded list-size is needed as one approaches the optimal rate of $1 - H_q(p)$. In other words, $L_{q,p}(\varepsilon) \to \infty$ as $\varepsilon \to 0$. This was shown for the binary case in [Bli86], and his result implicitly implies $L_{2,p}(\varepsilon) \geqslant \Omega(\log(1/\varepsilon))$ (see [Rud09] for an explicit derivation of this). For the q-ary case, $L_{q,p}(\varepsilon) = \omega_\varepsilon(1)$ was shown in [Bli05, Bli08]. In the language of list-decoding, the above-mentioned result on "almost-disjoint" sphere packing states that for large enough block lengths, a random code of rate $1 - H_q(p) - \varepsilon$ is $(p, \frac{1}{\varepsilon})$-list-decodable with high probability. In other words, $L_{q,p}(\varepsilon) \leqslant 1/\varepsilon$. This result appears in [Eli91] (and is based on a previous random coding argument for linear codes from [ZP82]). The result is explicitly stated in [Eli91] only for $q = 2$, but trivially extends for arbitrary alphabet size q. This result is also tight, in the sense that with high probability a *random* code of rate $1 - H_q(p) - \varepsilon$ is *not* $(p, c_{p,q}/\varepsilon)$-list-decodable w.h.p. for some constant $c_{p,q} > 0$ [Rud09].

An interesting question is to close the exponential gap in the lower and upper bounds on $L_{2,p}(\varepsilon)$, and more generally pin down the asymptotic behavior of $L_{q,p}(\varepsilon)$ for every q. The upper bound of $O(1/\varepsilon)$ is perhaps closer to the truth, and it is probably the lower bound that needs strengthening.

Context of this work. In this work, we address another fundamental combinatorial question concerning list-decodable codes, namely the behavior of $L_{q,p}(\varepsilon)$ when restricted to *linear codes*. For q a prime power, a q-ary linear code is simply a subspace of $\mathbb{F}_q^n$ ($\mathbb{F}_q$ being the field of size q).

Most of the well-studied and practically used codes are linear codes. Linear codes admit a succinct representation in terms of its basis (called generator matrix). This aids in finding and representing such codes efficiently, and as a result linear codes are often useful as "inner" codes in concatenated code constructions.

In a linear code, by translation invariance, the neighborhood of every codeword looks the same, and this is often a very useful symmetry property. For instance, this property was recently used in [GS09] to give a black-box conversion of linear list-decodable codes to codes achieving capacity against a worst-case additive channel (the linearity of the list-decodable code is crucial for this connection). Lastly, list-decodability of linear codes brings to the fore some intriguing questions on the interplay between the geometry of linear spaces and Hamming balls, and is therefore interesting in its own right. For these and several other reasons, it is desirable to achieve good trade-offs for list decoding via linear codes.

Since linear codes are a highly structured subclass of all codes, proving the existence of linear codes with list-decodability properties similar to general codes can be viewed as a strong "de-randomization" of the random coding argument used to construct good list-decodable codes. A derandomized family of codes called "pseudolinear codes" were put forth in [GI01] since linear codes were not known to have strong enough list-decoding properties. Indeed, prior to this work, the results known for linear codes were substantially weaker than for general codes (we discuss the details next). *Closing this gap is the main motivation behind this work.*

Status of list-decodability of linear codes. Zyablov and Pinsker proved that a random binary linear code of rate $1 - H(p) - \varepsilon$ is $(p, 2^{O(1/\varepsilon)})$-list-decodable with high probability [ZP82]. The proof extends in a straightforward way to linear codes over $\mathbb{F}_q$, giving list-size $q^{O(1/\varepsilon)}$ for rate

$1 - H_q(p) - \varepsilon$. Let us define $L^{\mathsf{lin}}_{q,p}(\varepsilon)$ to be the minimum integer L for which there is an infinite family of (p, L)-list-decodable linear codes over $\mathbb{F}_q$ of rate at least $1 - H_q(p) - \varepsilon$. The results of [ZP82] thus imply that $L^{\mathsf{lin}}_{q,p}(\varepsilon) \leqslant \exp(O_q(1/\varepsilon))$.

Note that this bound is *exponentially worse* than the $O(1/\varepsilon)$ bound known for general codes. In [Eli91], Elias mentions the following as the most obvious problem left open left by the random coding results: *Is the requirement of the much larger list size for linear codes inherent, or can one achieve list-size closer to the $O(1/\varepsilon)$ bound for general random codes?*

For the *binary* case, the *existence* of (p, L)-list-decodable linear codes of rate at least $1 - H(p) - 1/L$ is proven in [GHSZ02]. This implies that $L^{\mathsf{lin}}_{2,p} \leqslant 1/\varepsilon$. There are some results which obtain lower bounds on the rate for the case of small fixed list-size (at most 3) [Bli86, Bli97, WF94]; these bounds are complicated and not easily stated, and as noted in [Bli00], are weaker for the linear case for list-size as small as 5.

The proof in [GHSZ02] is based on a carefully designed potential function that quantifies list-decodability, and uses the "semi-random" method to successively pick good basis vectors for the code. The proof only guarantees that such binary linear codes exist with positive probability, and does not yield a high probability guarantee for the claimed list-decodability property. Further, the proof relies crucially on the binary alphabet and extending it to work for larger alphabets (or even the ternary case) has resisted all attempts. Thus, for $q > 2$, $L_{q,p}(\varepsilon) \leqslant \exp(O_q(1/\varepsilon))$ remained the best known upper bound on list-size. A high probability result for the binary case, and an upper bound of $L_{q,p}(\varepsilon) \leqslant O(1/\varepsilon)$ for $\mathbb{F}_q$-linear codes, were conjectured in [Gur04, Chap. 5].

Our contribution. In this work, we resolve the above open question concerning list-decodability of linear codes over *all* alphabets. In particular, we prove that $L^{\mathsf{lin}}_{q,p}(\varepsilon) \leqslant C_{q,p}/\varepsilon$ for a constant $C_{q,p} < \infty$. Up to constant factors, this matches the best known result for general, non-linear codes. Further, our result in fact shows that a random $\mathbb{F}_q$-linear code of rate $1-H_q(p)-\varepsilon$ is $(p, C_{p,q}/\varepsilon))$-list-decodable *with high probability*. This was not known even for the case $q = 2$. The high probability claim implies an efficient randomized Monte Carlo construction of such list-decodable codes.

We now briefly explain the difficulty in obtaining good bounds for list-decoding linear codes and how we circumvent it. This is just a high level description; see the next section for a more technical description of our proof method.

Let us recall the straightforward random coding method that shows the list-decodability of random (binary) codes. We pick a code $C \subseteq \{0,1\}^n$ by uniformly and independently picking $M = 2^{Rn}$ codewords. To prove it is (p, L)-list-decodable, we fix a center y and a subset S of $(L+1)$ codewords of C. Since these codewords are independent, the probability that all of them land in the ball of radius pn around y is at most $\left(\frac{2^{H(p)n}}{2^n}\right)^{L+1}$. A union bound over all 2^n choices of y and at most M^{L+1} choices of S shows that if $R \leqslant 1 - H(p) - 1/L$, the code fails to be (p, L)-list-decodable with probability at most $2^{-\Omega(n)}$.

Attempting a similar argument in the case of random linear codes, defined by a random linear map $A : \mathbb{F}_2^{Rn} \to \mathbb{F}_2^n$, faces several immediate obstacles. The 2^{Rn} codewords of a random linear code are not independent of one another; in fact the points of such a code are highly correlated and not even 3-wise independent (as $A(x + y) = Ax + Ay$). However, any $(L + 1)$ distinct codewords $Ax_1, Ax_2, \ldots, Ax_{L+1}$ must contain a subset of $\ell \geqslant \log_2(L+1)$ independent codewords, corresponding to a subset $\{x_{i_1}, \ldots, x_{i_\ell}\}$ of *linearly independent* message vectors. This lets one mimic the argument for the random code case with $\log_2(L + 1)$ playing the role of $L + 1$. However, as a result, it leads

to the exponentially worse list-size bounds.

To get a better result, we somehow need to control the "damage" caused by subsets of codewords of low rank. This is the crux of our new proof. Stated loosely and somewhat imprecisely, we prove a strong upper bound on the fraction of such low rank subsets, by proving that if we pick ℓ random vectors from the Hamming ball $B_n(0,p)$ (for some constant ℓ related to our target list-size L), it is rather unlikely that more than $\Theta(\ell)$ of the 2^ℓ vectors in their span will also belong to the ball $B_n(0,p)$. (See Theorem 3 for the precise statement.) This "limited correlation" between linear subspaces and Hamming balls is the main technical ingredient in our proof. It seems like a basic and powerful probabilistic fact that might find other applications. The argument also extends to linear codes over $\mathbb{F}_q$ after some adaptations.

2 Results and Methods

Our main result is that random linear codes in $\mathbb{F}_2^n$ of rate $1 - H(p) - \varepsilon$ can be list-decoded from p-fraction errors with list-size only $O(\frac{1}{\varepsilon})$. We also show the analogous result for random q-ary linear codes.

Theorem 2. *Let $p \in (0, 1/2)$. Then there exist constants $C_p, \delta > 0$, such that for all $\varepsilon > 0$ and all large enough integers n, letting $R = 1 - H(p) - \varepsilon$, if $\mathcal{C} \subseteq \mathbb{F}_2^n$ is a random linear code of rate R, then*

$$\Pr[\mathcal{C} \text{ is } (p, \tfrac{C_p}{\varepsilon})\text{-list-decodable}] > 1 - 2^{-\delta n}.$$

The proof begins by simplifying the problem to its combinatorial core. Specifically, we reduce the problem of studying the *list-decodability* of a random linear code of *linear* dimension to the problem of studying the *weight-distribution* of certain random linear codes of *constant* dimension. The next theorem analyzes the weight distribution of these constant-dimensional random linear codes. The notation $B_n(x,p)$ refers to the Hamming ball of radius pn centered at $x \in \mathbb{F}_2^n$.

Theorem 3 (Span of random points in $B_n(0,p)$)**.** *For every $p \in (0, 1/2)$, there is a constant $C > 0$, such that for all n large enough and all $\ell = o(\sqrt{n})$, if $X_1, \ldots, X_\ell$ are picked independently and uniformly at random from $B_n(0,p)$, then*

$$\Pr[|\mathrm{span}(\{X_1, \ldots, X_\ell\}) \cap B_n(0,p)| > C \cdot \ell] \leqslant 2^{-5n}.$$

We now give a brief sketch of the proof of Theorem 3. Index the elements of $\mathrm{span}(\{X_1, \ldots, X_\ell\})$ as follows: for $v \in \mathbb{F}_2^\ell$, let X_v denote the random vector $\sum_{i=1}^{\ell} v_i X_i$. Fix an arbitrary $S \subseteq \mathbb{F}_2^\ell$ of cardinality $C \cdot \ell$, and let us study the event E_S: that all the vectors $(X_v)_{v \in S}$ lie in $B_n(0,p)$. If none of the events E_S occur, we know that $|\mathrm{span}(\{X_1, \ldots, X_\ell\}) \cap B_n(0,p)| \leqslant C \cdot \ell$.

The key technical step is a Ramsey-theoretic lemma (Lemma 5, stated below) which says that large sets S automatically have the property that some translate of S contains a certain structured subset (which we call an "increasing chain"). This structured subset allows us to give strong upper bounds on the probability that all the vectors $(X_v)_{v \in S}$ lie in $B_n(0,p)$. Applying this to each $S \subseteq \mathbb{F}_2^\ell$ of cardinality $C\ell$ and taking a union bound gives Theorem 3.

To state the Ramsey-theoretic lemma (Lemma 5), we first define increasing chains. For a vector $v \in \mathbb{F}_2^\ell$, the *support* of v, denoted $\mathrm{supp}(v)$, is defined to be the set of its nonzero coordinates.

Definition 4. *A sequence of vectors $v_1, \ldots, v_d \in \mathbb{F}_2^\ell$ is called an c-increasing chain of length d, if for all $j \in [d]$,*

$$\left| \mathrm{supp}(v_j) \setminus \left(\bigcup_{i=1}^{j-1} \mathrm{supp}(v_i) \right) \right| \geqslant c.$$

We now state the Ramsey-theoretic lemma that plays the central role in Theorem 3. The proof appears in Section 5, where it is proved using the Sauer-Shelah lemma.

Lemma 5. *For all positive integers c, ℓ and $L \leqslant 2^\ell$, the following holds. For every $S \subseteq \mathbb{F}_2^\ell$ with $|S| = L$, there is a $w \in \mathbb{F}_2^\ell$ such that $S + w$ has an c-increasing chain of length at least $\frac{1}{c}(\log \frac{L}{2}) - (1 - \frac{1}{c})(\log \ell)$.*

2.1 Larger alphabet

Due to their geometric nature, our arguments generalize to the case of q-ary alphabet (for arbitrary constant q) quite easily. Below we state our main theorem for the case of q-ary alphabet.

Theorem 6. *Let q be a prime power and let $p \in (0, 1-1/q)$. Then there exist constants $C_{p,q}, \delta > 0$, such that for all $\varepsilon > 0$, letting $R = 1 - H_q(p) - \varepsilon$, if $\mathcal{C} \subseteq \mathbb{F}_q^n$ is a random linear code of rate R, then*

$$\Pr[\mathcal{C} \text{ is } (p, \tfrac{C_{p,q}}{\varepsilon})\text{-list-decodable}] > 1 - 2^{-\delta n}.$$

The proof of Theorem 6 has the same basic outline as the proof of Theorem 2. In particular, it proceeds via a q-ary analog of Theorem 3. The only notable deviation occurs in the proof of the q-ary analog of Lemma 5. The traditional generalization of the Sauer-Shelah lemma to larger alphabets turns out to be unsuitable for this purpose. Instead, we formulate and prove a non-standard generalization of the Sauer-Shelah lemma for the larger alphabet case which is more appropriate for this situation. Details appear in Section 6.

3 Proof of Theorem 2

Let us start by restating our main theorem.

Theorem 2 (restated) *Let $p \in (0, 1/2)$. Then there exist constants $C_p, \delta > 0$, such that for all $\varepsilon > 0$ and all large enough integers n, letting $R = 1 - H(p) - \varepsilon$, if $\mathcal{C} \subseteq \mathbb{F}_2^n$ is a random linear code of rate R, then*

$$\Pr[\mathcal{C} \text{ is } (p, \tfrac{C_p}{\varepsilon})\text{-list-decodable}] > 1 - 2^{-\delta n}.$$

Proof. Pick $C_p = 4C$, where C is the constant from Theorem 3. Pick $\delta = 1$. Take $L = \frac{C_p}{\varepsilon}$.

Let $\mathcal{C}$ be a random Rn dimensional linear subspace of $\mathbb{F}_2^n$. We want to show that

$$\Pr_{\mathcal{C}}[\exists x \in \mathbb{F}_2^n \text{ s.t. } |B_n(x, p) \cap \mathcal{C}| > L] < 2^{-\delta n}. \tag{1}$$

Let $x \in \mathbb{F}_2^n$ be picked uniformly at random. We will work towards Equation (1) by studying the following quantity.

$$\Delta \stackrel{\text{def}}{=} \Pr_{\mathcal{C},x}[|B_n(x,p) \cap \mathcal{C}| > L].$$

Note that to prove Equation (1), it suffices to show that[1]

$$\Delta < 2^{-\delta n} \cdot 2^{-n}.$$

Now for each $\ell \in [\log(L+1), L+1]$, let $\mathcal{F}_\ell$ be the set of all $(v_1, \ldots, v_\ell) \in B_n(0,p)^\ell$ such that $v_1, \ldots, v_\ell$ are linearly independent and $|\text{span}(v_1, \ldots, v_\ell) \cap B_n(0,p)^\ell| > L$. Let $\mathcal{F} = \bigcup_{\ell=\log(L+1)}^{L+1} \mathcal{F}_\ell$

For each $\mathbf{v} = (v_1, \ldots, v_\ell) \in \mathcal{F}$, let $\{\mathbf{v}\}$ denote the set $\{v_1, \ldots, v_\ell\}$.

We now bound Δ. Notice that if $|B_n(x,p) \cap \mathcal{C}| > L$, then there must be some $\mathbf{v} \in \mathcal{F}$ for which $B_n(x,p) \cap \mathcal{C} \supseteq x + \{\mathbf{v}\}$. Indeed, we can simply take $\mathbf{v}$ to be a maximal linearly independent subset of $(B_n(x,p) \cap \mathcal{C}) + x$ if this set has size at most $L+1$, and any linearly independent subset of $(B_n(x,p) \cap \mathcal{C}) + x$ of size $L+1$ otherwise.

Therefore, by the union bound,

$$\begin{aligned}
\Delta &\leqslant \sum_{\mathbf{v} \in \mathcal{F}} \Pr_{\mathcal{C},x}[B_n(x,p) \cap \mathcal{C} \supseteq x + \{\mathbf{v}\}] && (2)\\
&= \sum_{\mathbf{v} \in \mathcal{F}} \Pr_{\mathcal{C},x}[B_n(0,p) \cap (\mathcal{C} + x) \supseteq \{\mathbf{v}\}] && (3)\\
&\leqslant \sum_{\mathbf{v} \in \mathcal{F}} \Pr_{\mathcal{C},x}[B_n(0,p) \cap (\mathcal{C} + \{0,x\}) \supseteq \{\mathbf{v}\}] && (4)\\
&= \sum_{\mathbf{v} \in \mathcal{F}} \Pr_{\mathcal{C}^*}[B_n(0,p) \cap \mathcal{C}^* \supseteq \{\mathbf{v}\}], && (5)
\end{aligned}$$

where $\mathcal{C}^*$ is the code $\mathcal{C} + \{0,x\}$ which is a random $Rn+1$ dimensional subspace.

The last probability can be bounded as follows. By the linear independence of $v_1, \ldots, v_\ell$, the probability that $v_j \in \mathcal{C}^*$ conditioned on $\{v_1, \ldots, v_{j-1}\} \subseteq \mathcal{C}^*$ is precisely the probability that a given point in a $n+1-j$ dimensional space lies in a $Rn+1-j$ dimensional subspace, and hence this conditional probability is exactly 2^{Rn+1-n}. We can hence conclude that

$$\Pr_{\mathcal{C}^*}[\mathcal{C}^* \supseteq \{\mathbf{v}\}] = \left(\frac{2^{Rn+1}}{2^n}\right)^\ell. \quad (6)$$

[1] We could even replace the 2^{-n} by $2^{-(1-R)n}$. Indeed, for every $\mathcal{C}$ for which there is a "bad" x, we know that there are 2^{Rn} "bad" x's (the translates of x by $\mathcal{C}$).

Putting together Equations (5) and (6), we have

$$\begin{aligned}\Delta &\leqslant \sum_{\mathbf{v}\in\mathcal{F}} \Pr_{\mathcal{C}^*}[B_n(0,p)\cap\mathcal{C}^* \supseteq \{\mathbf{v}\}] \leqslant \sum_{\ell=\log(L+1)}^{L+1} \sum_{\mathbf{v}\in\mathcal{F}_\ell} \Pr_{\mathcal{C}^*}[\mathcal{C}^* \supseteq \{\mathbf{v}\}] \\ &\leqslant \sum_{\ell=\log(L+1)}^{L+1} \sum_{\mathbf{v}\in\mathcal{F}_\ell} \left(\frac{2^{Rn+1}}{2^n}\right)^\ell \leqslant \sum_{\ell=\log(L+1)}^{L+1} |\mathcal{F}_\ell| \cdot \left(\frac{2^{Rn+1}}{2^n}\right)^\ell\end{aligned}$$

We now obtain an upper bound on $|\mathcal{F}_\ell|$. We have two cases depending on the size of ℓ.

- **Case 1:** $\ell < 4/\varepsilon$. In this case, we notice that $\frac{|\mathcal{F}_\ell|}{|B_n(0,p)|^\ell}$ is a lower bound on the probability that ℓ points $X_1,\ldots,X_\ell$ chosen uniformly at random from $B_n(0,p)$ have $|\mathrm{span}(\{X_1,\ldots,X_\ell\})\cap B_n(0,p)| > L$. Since $L > C\cdot\ell$, Theorem 3 tells us that this probability is bounded from above by 2^{-5n}. Thus, in this case $|\mathcal{F}_\ell| \leqslant |B_n(0,p)|^\ell 2^{-5n} \leqslant 2^{n\ell H(p)}\cdot 2^{-5n}$.
- **Case 2:** $\ell \geqslant 4/\varepsilon$. In this case, we have the trivial bound of $|\mathcal{F}_\ell| \leqslant |B_n(0,p)|^\ell \leqslant 2^{n\ell H(p)}$.

Thus, we may bound Δ by:

$$\begin{aligned}\Delta &\leqslant \sum_{\ell=\log L}^{\lfloor 4/\varepsilon\rfloor} |\mathcal{F}_\ell| \cdot \left(\frac{2^{Rn+1}}{2^n}\right)^\ell + \sum_{\ell=\lceil 4/\varepsilon\rceil}^{L} |\mathcal{F}_\ell| \cdot \left(\frac{2^{Rn+1}}{2^n}\right)^\ell \\ &\leqslant \sum_{\ell=\log L}^{\lfloor 4/\varepsilon\rfloor} 2^{n\ell H(p)} 2^{-5n} \left(\frac{2^{Rn+1}}{2^n}\right)^\ell + \sum_{\ell=\lceil 4/\varepsilon\rceil}^{L} 2^{n\ell H(p)} \left(\frac{2^{Rn+1}}{2^n}\right)^\ell \\ &\leqslant 2^{-5n}\cdot 4/\varepsilon + L\cdot 2^{-(\varepsilon n)\cdot(4/\varepsilon)} \\ &\leqslant 2^{-\delta n}\cdot 2^{-n}\end{aligned}$$

as desired. □

4 Proof of Theorem 3

In this section, we prove Theorem 3 which bounds the probability that the span of ℓ random points in $B_n(0,p)$ intersects $B_n(0,p)$ in more than $C\cdot\ell$ points, for some large constant C. We use the following simple fact.

Lemma 7. *For every $p\in(0,1/2)$, there is a $\delta_p > 0$ such that for all large enough integers n and every $x\in\mathbb{F}_2^n$, the probability that two uniform independent samples w_1, w_2 from $B_n(0,p)$ are such that $w_1+w_2\in B_n(x,p)$ is at most $2^{-\delta_p n}$.*

Sketch of proof. The point w_1+w_2 is essentially a random point in $B_n(0,2p-2p^2)$. The probability that it lies in the smaller ball $B_n(x,p)$ is easily seen to be maximal when $x=0$ and is then exponentially small. □

Theorem 3 (restated) *For every $p \in (0, 1/2)$, there is a constant $C > 0$, such that for all n large enough and all $\ell = o(\sqrt{n})$, if $X_1, \ldots, X_\ell$ are picked independently and uniformly at random from $B_n(0, p)$, then*

$$\Pr[|\mathrm{span}(\{X_1, \ldots, X_\ell\}) \cap B_n(0, p)| > C \cdot \ell] \leqslant 2^{-5n}.$$

Proof. Set $L = C \cdot \ell$ and let $c = 2$. Let $\delta_p > 0$ be the constant given by Lemma 7. Let

$$d = \left\lfloor \frac{1}{c} \log \frac{L}{2} - \left(1 - \frac{1}{c}\right) \log \ell \right\rfloor \geqslant \frac{1}{2} \log \frac{L}{2\ell} - 1 = \frac{1}{2} \log \frac{C}{8} .$$

For a vector $u \in \mathbb{F}_2^\ell$, let X_u denote the random variable $\sum_i u_i X_i$.

We begin with a claim which bounds the probability of a particular collection of linear combinations of the X_i all lying within $B_n(0, p)$. At the heart of this claim lies the Ramsey-theoretic Lemma 5.

Claim 8. *For each $S \subseteq \mathbb{F}_2^\ell$ with $|S| = L + 1$,*

$$\Pr[\forall v \in S, X_v \in B_n(0, p)] < 2^n \cdot 2^{-\delta_p d n}. \tag{7}$$

Proof. Let w and $v_1, \ldots, v_d \in S$ be as given by Lemma 5. That is, $v_1 + w, v_2 + w, \cdots, v_d + w$ is an c-increasing sequence. Then,

$$\begin{aligned}
\Pr[\forall v \in S, X_v \in B_n(0, p)] &\leqslant \Pr[\forall j \in [d], X_{v_j} \in B_n(0, p)] &(8)\\
&= \Pr[\forall j \in [d], X_{v_j} + X_w \in B_n(X_w, p)] &(9)\\
&= \Pr[\forall j \in [d], X_{v_j + w} \in B_n(X_w, p)] &(10)
\end{aligned}$$

We now bound the probability that there exists $y \in \mathbb{F}_2^n$ such that for all $j \in [d]$, $X_{v_j+w} \in B_n(y, p)$. Fix $y \in \mathbb{F}_2^n$. We have:

$$\begin{aligned}
\Pr[\forall j \in [d], X_{v_j+w} \in B_n(y, p)] &\leqslant \prod_{j=1}^{d} \Pr\left[X_{v_j+w} \in B_n(y, p) \mid (X_t : t \in \left(\bigcup_{i=1}^{j-1} \mathrm{supp}(v_i + w)\right)\right] &(11)\\
&\leqslant \left(2^{-\delta_p n}\right)^d. &(12)
\end{aligned}$$

The last inequality follows from applying Lemma 7 with w_1 and w_2 being vectors X_{i_1} and X_{i_2}, where i_1, i_2 are two distinct elements of $\mathrm{supp}(v_j+w) \setminus \left(\bigcup_{i=1}^{j-1} \mathrm{supp}(v_i + w)\right)$, and $x = y + \sum_{k \in [\ell], k \notin \{i_1, i_2\}} (v_j + w)_k X_k$. Taking a union bound of Equation (12) over all $y \in \mathbb{F}_2^n$, we see that

$$\Pr[\exists y \in F_2^n \text{ s.t. } \forall j \in [d], X_{v_j+w} \in B_n(y, p)] \leqslant 2^n \cdot 2^{-\delta_p n d}.$$

Combining this with Equation (10) completes the proof of the claim. □

Given this claim, we now bound the probability that more than L elements of $\mathrm{span}(\{X_1, \ldots, X_\ell\})$ lie inside $B_n(0, p)$. This event occurs if and only if for some set $S \subseteq \mathbb{F}_2^\ell$ with $|S| = L + 1$, it is the case that $\forall v \in S$, $X_v \in B_n(0, p)$. Taking a union bound of (7) over all such S, we see that the probability that there exists some $S \subseteq \mathbb{F}_2^\ell$ with $|S| = L + 1$ such that $\forall v \in S, X_v \in B_n(0, p)$ is at most $2^{\ell(L+1)} \cdot 2^n \cdot 2^{-\delta_p d n}$. Taking C to be a large enough constant so that $d \geqslant \frac{1}{2} \log \frac{C}{8} > \frac{12}{\delta_p}$, the theorem follows. □

5 Proof of Lemma 5

In this section, we will prove Lemma 5, which finds a large c-increasing chain in some translate of any large enough set $S \subseteq \mathbb{F}_2^\ell$.

We will use the Sauer-Shelah Lemma.

Lemma 9 (Sauer-Shelah [Sau72, She72])**.** *For all integers ℓ, c, and for any set $S \subseteq \{0,1\}^\ell$, if $|S| > 2\ell^{c-1}$, then there exists some set of coordinates $U \subseteq [\ell]$ with $|U| = c$ such that $\{v|_U \mid v \in S\} = \{0,1\}^U$.*

Lemma 5 (restated) *For all positive integers c, ℓ and $L \leqslant 2^\ell$, the following holds. For every $S \subseteq \mathbb{F}_2^\ell$ with $|S| = L$, there is a $w \in \mathbb{F}_2^\ell$ such that $S + w$ has an c-increasing chain of length at least $\frac{1}{c}(\log \frac{L}{2}) - (1 - \frac{1}{c})(\log \ell)$.*

Proof. We prove this by induction on ℓ. The claim holds trivially for $\ell \leqslant c$, so assume $\ell > c$.

If $L \leqslant 2\ell^{c-1}$, then again the lemma holds trivially. Otherwise, by the Sauer-Shelah lemma, we get a set U of c coordinates such that for each $u \in \mathbb{F}_2^U$, there is some $v \in S$ such that $v|_U = u$. We will represent elements of $\mathbb{F}_2^\ell$ in the form (u, v') where $u \in \mathbb{F}_2^U$ and $v' \in \mathbb{F}_2^{[\ell]\setminus U}$.

Let $u_0 \in \mathbb{F}_2^U$ be a vector such that $|\{v \in S \mid v|_U = u_0\}|$ is at least $L/2^c$ (we know that such a u exists by averaging). Let $S' \subseteq \mathbb{F}_2^{[\ell]\setminus U}$ be given by $S' = \{v|_{[\ell]\setminus U} \mid v|_U = u\}$. By choice of u, we have $|S'| \geqslant L/2^c$.

By the induction hypothesis, there exist $w' \in \mathbb{F}_2^{\ell-c}$ and $v'_1, \ldots, v'_{d'} \in S'$ such that for each $j \in [d']$,

$$\left|\operatorname{supp}(v'_j + w') \setminus \left(\bigcup_{i=1}^{j-1} \operatorname{supp}(v'_i + w')\right)\right| \geqslant c.$$

for $d' \geqslant \frac{1}{c}\log(L/2^{c+1}) - (1 - \frac{1}{c})\log(\ell - c)$.

Let $d = d' + 1$. Note that $d \geqslant \frac{1}{c}\log(L/2) - (1 - \frac{1}{c})\log(\ell - c) \geqslant \frac{1}{c}\log(L/2) - (1 - \frac{1}{c})\log \ell$. For $i \in [d']$, let $v_i = (u_0, v'_i) \in \mathbb{F}_2^\ell$. Let v_d be any vector in S with $(v_d)|_U = \neg u_0$, the bitwise complement of u_0. Let $w = (u_0, w')$. We claim that w and $v_1, \ldots, v_d$ satisfy the desired properties.

Indeed, for each $j \in [d']$, we have

$$\left|\operatorname{supp}(v_j + w) \setminus \left(\bigcup_{i=1}^{j-1} \operatorname{supp}(v_i + w)\right)\right| = \left|\operatorname{supp}(v'_j + w') \setminus \left(\bigcup_{i=1}^{j-1} \operatorname{supp}(v'_i + w')\right)\right| \geqslant c.$$

Also

$$\left|\operatorname{supp}(v_d + w) \setminus \left(\bigcup_{i=1}^{d-1} \operatorname{supp}(v_i + w)\right)\right| \geqslant |\operatorname{supp}(v_d + w) \setminus ([\ell] \setminus U)| = |U| = c.$$

Thus for all $j \in [d]$, we have $\left|\operatorname{supp}(v_j + w) \setminus \left(\bigcup_{i=1}^{j-1} \operatorname{supp}(v_i + w)\right)\right| \geqslant c$, as desired. ∎

6 Larger alphabets

As mentioned in the introduction the case of q-ary alphabet is nearly identical to the case of binary alphabet. We only highlight the differences. As before, the crux turns out to be the problem of studying the weight distribution of certain random constant-dimensional codes.

Theorem 10 (q-ary span of random points in $B_n^q(0,p)$)**.** *For every prime-power q and every $p \in (0, 1-1/q)$, there is a constant $C_q > 0$, such that for all n large enough and all $\ell = o(\sqrt{n})$, if $X_1, \ldots, X_\ell$ are picked independently and uniformly at random from $B_n^q(0,p)$, then*

$$\Pr[|\mathrm{span}(\{X_1, \ldots, X_\ell\}) \cap B_n^q(0,p)| > C_q \cdot \ell] \leqslant q^{-5n}.$$

The proof of Theorem 10 proceeds as before, by bounding the probability via a large c-increasing chain. The c-increasing chain itself is found in an analog of Lemma 5 for q-ary alphabet. We first need a definition.

Definition 11. *A sequence of vectors $v_1, \ldots, v_d \in [q]^\ell$ is called an c-increasing chain of length d, if for all $j \in [d]$,*

$$\left| \mathrm{supp}(v_j) \setminus \left(\bigcup_{i=1}^{j-1} \mathrm{supp}(v_i) \right) \right| \geqslant c.$$

Now we have the following lemma.

Lemma 12 (q-ary increasing chains Ramsey)**.** *For every prime power q, and all positive integers c, ℓ and $L \leqslant q^\ell$, the following holds. For every $S \subseteq \mathbb{F}_q^\ell$ with $|S| = L$, there is a $w \in \mathbb{F}_q^\ell$ such that $S + w$ has an c-increasing chain of length at least $\frac{1}{c}\log_q\left(\frac{L}{2}\right) - (1 - \frac{1}{c})\log_q((q-1)\ell)$.*

The proof of Lemma 12 needs a non-standard generalization of the Sauer-Shelah lemma to larger alphabet described in the next section.

6.1 A q-ary Sauer-Shelah lemma

The traditional generalization of the Sauer-Shelah lemma to large alphabets is the Karpovsky-Milman lemma [KM78], which roughly states that given $S \subseteq [q]^\ell$ of cardinality at least $(q-1)^l l^{c-1}$, there is a set U of c coordinates such that for every $u \in [q]^U$, there is some $v \in S$ such that the restriction $v|_U$ equals u. Applying this lemma in our context, once $q > 2$, requires us to have a set $S > 2^\ell$, which turns out to lead to exponential list size bounds. Fortunately, the actual property needed for us is slightly different. We want a bound B (ideally polynomial in ℓ) such that for any set $S \subseteq [q]^\ell$ of cardinality at least B, there is a set U of c coordinates such that for every $u \in [q]^U$, there is some $v \in S$ such that the restriction $v|_U$ *differs from u in every coordinate of U*. It turns out that this weakened requirement admits polynomial-sized B.

We state and prove this generalization of the Sauer-Shelah lemma below.

Lemma 13 (q-ary Sauer-Shelah)**.** *For all integers q, ℓ, c, for any set $S \subseteq [q]^\ell$, if $|S| > 2 \cdot ((q-1) \cdot \ell)^{c-1}$, then there exists some set of coordinates $U \subseteq [\ell]$ with $|U| = c$ such that for every $u \in [q]^U$, there exists some $v \in S$ such that u and $v|_U$ differ in every coordinate.*

Proof. We prove this by induction on ℓ and c. If $c = 1$, then $|S| > 2$ and the result holds by letting U equal any coordinate on which not all elements of S agree.

Now assume $c > 1$. Represent an element x of $[q]^\ell$ as a pair (y, b), where $y \in [q]^{\ell-1}$ consists of the first $\ell - 1$ coordinates of x and $b \in [q]$ is the last coordinate of x.

Consider the following subsets of $[q]^{\ell-1}$.

$$S_1 = \{y \in [q]^{\ell-1} \mid \text{ for at least 1 value of } b \in [q],\ (y, b) \in S\}.$$

$$S_2 = \{y \in [q]^{\ell-1} \mid \text{ for at least 2 values of } b \in [q],\ (y, b) \in S\}.$$

Note that $|S| \leqslant (|S_1| - |S_2|) + q|S_2| = |S_1| + (q-1)|S_2|$. By assumption,

$$|S| > 2 \cdot ((q-1) \cdot \ell)^{c-1} \geqslant 2 \cdot ((q-1) \cdot (\ell-1))^{c-1} + (q-1)\left(2 \cdot ((q-1) \cdot (\ell-1))^{c-2}\right),$$

(using the elementary inequality $\ell^{c-1} \geqslant (\ell-1)^{c-1} + (\ell-1)^{c-2}$). Thus, either $|S_1| > 2 \cdot ((q-1) \cdot (\ell-1))^{c-1}$, or else $|S_2| > 2 \cdot ((q-1) \cdot (\ell-1))^{c-2}$.

We now prove the desired claim in each of these cases.

Case 1: $|S_1| > 2 \cdot ((q-1) \cdot (\ell-1))^{c-1}$. In this case, we can apply the induction hypothesis to S_1 with parameters $\ell - 1$ and c, and get a subset of U of $[\ell - 1]$ of cardinality c. Then the set U has the desired property.

Case 2: $|S_2| > 2 \cdot ((q-1) \cdot (\ell-1))^{c-2}$. In this case, we apply the induction hypothesis to S_2 with parameters $\ell - 1$ and $c - 1$, and get a subset U of $[\ell - 1]$ of cardinality $c - 1$. Then the set $U \cup \{\ell\}$ has the desired property. Indeed, take any vector $u \in [q]^{U \cup \{\ell\}}$. Let $u' = u|_U$. By the induction hypothesis, we know that there is a $v \in S_2$ such that $v|_U$ differs from u' in every coordinate of U. Now we know that there are at least two $b \in [q]$ such that $(v, b) \in S$. At least one of these b will be such that (v, b) differs from u in every coordinate of $U \cup \{\ell\}$, as desired. □

In the next section, we use the above lemma to prove the Ramsey-theoretic q-ary increasing chain claim (Lemma 12).

7 Proof of q-ary increasing chain lemma

In this section, we prove Lemma 12, which we restate below for convenience.

Lemma 12 (restated) *For every prime power q, and all positive integers c, ℓ and $L \leqslant q^\ell$, the following holds. For every $S \subseteq \mathbb{F}_q^\ell$ with $|S| = L$, there is a $w \in \mathbb{F}_q^\ell$ such that $S + w$ has an c-increasing chain of length at least $\frac{1}{c}\log_q\left(\frac{L}{2}\right) - \left(1 - \frac{1}{c}\right)\log_q((q-1)\ell)$.*

Proof. We prove this by induction on ℓ. The claim holds trivially for $\ell \leqslant c$, so assume $\ell > c$.

If $L \leqslant 2((q-1) \cdot \ell)^{c-1}$, then again the lemma holds trivially. Otherwise, by Lemma 13 we get a set U of c coordinates such that for each $u \in \mathbb{F}_q^U$, there is some $v \in S$ such that $v|_U$ differs from u in every coordinate. We will represent elements of $\mathbb{F}_q^\ell$ in the form (u, v') where $u \in \mathbb{F}_q^U$ and $v' \in \mathbb{F}_q^{[\ell] \setminus U}$.

Let $u_0 \in \mathbb{F}_q^U$ be a vector such that $|\{v \in S \mid v|_U = u_0\}|$ is at least L/q^c (we know that such a u exists by averaging). Let $S' \subseteq \mathbb{F}_q^{[\ell]\setminus U}$ be given by $S' = \{v|_{[\ell]\setminus U} \mid v|_U = u\}$. By choice of u, we have $|S'| \geqslant L/q^c$.

By the induction hypothesis, for

$$d' \geqslant \frac{1}{c}\log\Big(\frac{L}{2q^c}\Big) - \Big(1 - \frac{1}{c}\Big)\log((q-1)(\ell-c)) \ ,$$

there exist $w' \in \mathbb{F}_q^{\ell-c}$ and $v'_1, \ldots, v'_{d'} \in S'$ such that for each $j \in [d']$,

$$\Big|\mathrm{supp}(v'_j + w') \setminus \Big(\bigcup_{i=1}^{j-1} \mathrm{supp}(v'_i + w')\Big)\Big| \geqslant c.$$

Let $d = d' + 1$. Note that

$$d \geqslant \frac{1}{c}\log_q\Big(\frac{L}{2}\Big) - \Big(1 - \frac{1}{c}\Big)\log_q((q-1)\ell) \ .$$

For $i \in [d']$, let $v_i = (u_0, v'_i) \in \mathbb{F}_q^\ell$. Let v_d be any vector in S where $(v_d)|_U$ differs from u_0 in every coordinate of U. Let $w = (-u_0, w')$. We claim that w and $v_1, \ldots, v_d$ satisfy the desired properties. Indeed, for each $j \in [d']$, we have

$$\Big|\mathrm{supp}(v_j + w) \setminus \Big(\bigcup_{i=1}^{j-1} \mathrm{supp}(v_i + w)\Big)\Big| = \Big|\mathrm{supp}(v'_j + w') \setminus \Big(\bigcup_{i=1}^{j-1} \mathrm{supp}(v'_i + w')\Big)\Big| \geqslant c.$$

Also,

$$\Big|\mathrm{supp}(v_d + w) \setminus \Big(\bigcup_{i=1}^{d-1} \mathrm{supp}(v_i + w)\Big)\Big| \geqslant \Big|\mathrm{supp}(v_d + w) \setminus ([\ell] \setminus U)\Big| = |U| = c.$$

Thus for all $j \in [d]$, we have

$$\Big|\mathrm{supp}(v_j + w) \setminus \Big(\bigcup_{i=1}^{j-1} \mathrm{supp}(v_i + w)\Big)\Big| \geqslant c,$$

as desired. □

Given Lemma 12, the proof of Theorem 10 is virtually identical to the proof of its binary analog Theorem 3. Theorem 6 can then be proved (using Theorem 10) in the same manner as Theorem 2 was proved.

Acknowledgements

Some of this work was done when we were all participating in the Dagstuhl seminar 09441 on constraint satisfaction. We thank the organizers of the seminar for inviting us, and Schloss Dagstuhl for the wonderful hospitality.

References

[Bli86] Volodia M. Blinovsky. Bounds for codes in the case of list decoding of finite volume. *Problems of Information Transmission*, 22(1):7–19, 1986. 3, 4

[Bli97] Volodia M. Blinovsky. *Asymptotic Combinatorial Coding Theory*. Kluwer Academic Publishers, Boston, 1997. 4

[Bli00] Volodia M. Blinovsky. Lower bound for the linear multiple packing of the binary hamming space. *Journal of Combinatorial Theory, Series A*, 92(1):95–101, 2000. 4

[Bli05] Volodia M. Blinovsky. Code bounds for multiple packings over a nonbinary finite alphabet. *Problems of Information Transmission*, 41(1):23–32, 2005. 3

[Bli08] Volodia M. Blinovsky. On the convexity of one coding-theory function. *Problems of Information Transmission*, 44(1):34–39, 2008. 3

[Eli57] Peter Elias. List decoding for noisy channels. *Technical Report 335, Research Laboratory of Electronics, MIT*, 1957. 2

[Eli91] Peter Elias. Error-correcting codes for list decoding. *IEEE Transactions on Information Theory*, 37:5–12, 1991. 3, 4

[GHSZ02] Venkatesan Guruswami, Johan Håstad, Madhu Sudan, and David Zuckerman. Combinatorial bounds for list decoding. *IEEE Transactions on Information Theory*, 48(5):1021–1035, 2002. 4

[GI01] Venkatesan Guruswami and Piotr Indyk. Expander-based constructions of efficiently decodable codes. In *Proceedings of the 42nd Annual IEEE Symposium on Foundations of Computer Science*, pages 658–667, 2001. 3

[GS09] Venkatesan Guruswami and Adam D. Smith. Explicit capacity-achieving codes for worst-case additive errors. Preprint, arxiv:0912.0965 [cs.IT], 2009. 3

[Gur04] Venkatesan Guruswami. *List decoding of error-correcting codes*. Number 3282 in Lecture Notes in Computer Science. Springer, 2004. 4

[KM78] M. G. Karpovsky and V. D. Milman. Coordinate density of sets of vectors. *Discrete Math.*, 24(2):177–184, 1978. 11

[Rud09] Atri Rudra. Limits to list decoding random codes. In Hung Q. Ngo, editor, *COCOON*, volume 5609 of *Lecture Notes in Computer Science*, pages 27–36. Springer, 2009. 3

[Sau72] N. Sauer. On the density of families of sets. *J. Combinatorial Theory Ser. A*, 13:145–147, 1972. 10

[She72] Saharon Shelah. A combinatorial problem; stability and order for models and theories in infinitary languages. *Pacific J. Math.*, 41:247–261, 1972. 10

[WF94] Victor K. Wei and Gui-Liang Feng. Improved lower bounds on the sizes of error-correcting codes for list decoding. *IEEE Transactions on Information Theory*, 40(2):559–563, 1994. 4

[Woz58] John M. Wozencraft. List Decoding. *Quarterly Progress Report, Research Laboratory of Electronics, MIT*, 48:90–95, 1958. 2

[ZP82] Victor V. Zyablov and Mark S. Pinsker. List cascade decoding. *Problems of Information Transmission*, 17(4):29–34, 1981 (in Russian); pp. 236-240 (in English), 1982. 3